Image Retrieval Based on Text and Visual Content Using Neural Networks

Senior Undergraduate Thesis
Diego A. Castro
dcastro@dc.uba.ar
Director: Leticia Seijas
lseijas@dc.uba.ar
Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires, Argentina

Keywords: image retrieval, Self-Organizing Maps (SOM), Content-Based Image Retrieval (CBIR), Text-Based Image Retrieval (TBIR), color histogram, ParBSOM, scoring function, Envision

1 Introduction

In the last few years there has been a dramatic increase in the visual information available. Images generated from satellites, surveillance cameras and even digital cameras produce a huge amount of information that gradually becomes more difficult to handle. In the image retrieval area (VIR), images are typically described by their textual content (TBIR) or by their visual features (CBIR). However, these approaches still present many problems. While in TBIR using natural language can lead to subjective and ambiguous descriptions, CBIR uses low-level features and can regard images as similar when they are semantically different -a problem known as *semantic gap* [1]-.

Recently, the hybrid approach was introduced. It combines both characteristics to improve the benefits of using text and visual content separately. CBIR nowadays is still far from being as well-matured as TBIR since it presents many challenges such as defining suitable descriptors and index structures.

In this work we first focus on investigating techniques related to CBIR. We study one of the most popular image descriptors in the area: the *color histograms* [2]. We also investigate how *Self-Organizing Maps* (SOM) [3] can be used as an index in CBIR. SOM is an interesting alternative as it allows us to work with high-dimensional descriptors (typical case in CBIR). We propose a *scoring function* for images which eliminates irrelevant images from the results list and we also introduce a new SOM model that improves training and retrieval times (*ParBSOM*).

In order to evaluate the performance of the studied methods, we base our experiments on image databases which are used in many works of the area or in events like ImageCLEF. Specifically, we use ZuBuD [4], UCID [5], UK Bench [6] and ImageCLEFphoto 2007 [7]. We also work with typical retrieval metrics such as Precision, Recall, F-Measure and MAP.

In addition, we study how these techniques can be applied to the hybrid approach and provide computational results to assess their performance. Finally, we develop a research system known as *Envision*, which implements all the studied methods and was designed with extensibility and flexibility in mind.

2 Methods

Color is one of the most intuitive features of an image which explains why *color histograms* are among the most widely used features in the area. The color histogram for an image is constructed by counting the number of pixels of each color. This descriptor can work with different color spaces such as RGB or HSV. In many works, HSV has been used as it is perceptually more uniform than the popular RGB [8][1]. To work with color histograms, a distance measure must be defined to determine how close images are. The L1 distance measure showed improved results in several works [9].

Typically, in TBIR a scoring function is defined and used to retrieve only meaningful results. However, this topic has been neglected in CBIR. In order to eliminate irrelevant images from the results list, we propose a scoring function that allows us to define a thresh-

[1] In our experiments, using HSV improved results (between 15 and 40% in all image databases)

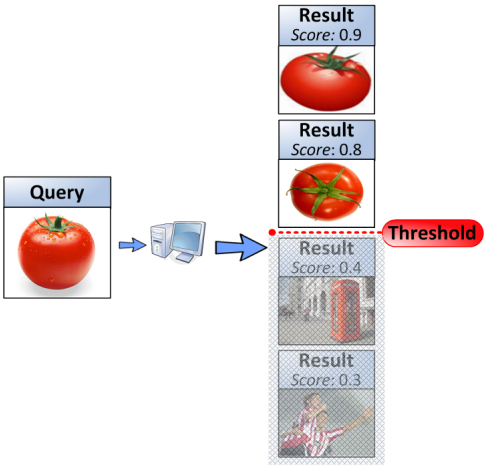

Figure 1: Threshold to eliminate irrelevant images during retrieval

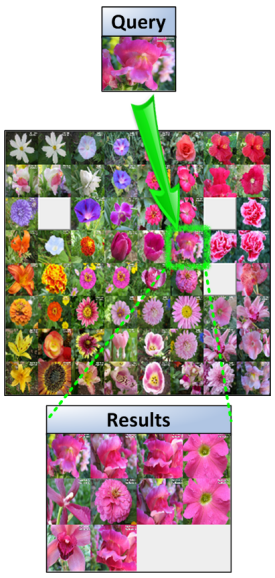

Figure 2: Retrieving images from a trained SOM

old (between 0 and 1) and filter those images below it (Fig. 1). We formally demonstrate that this function is valid when using color histograms together with the L1 distance measure.

One of the main problems faced in CBIR is that image descriptors are usually high-dimensional and current techniques such as R-Trees [10] or KD-Trees [10] are not scalable for dimensions higher than 20. In this context, SOM is an interesting alternative as it allows us to work with high-dimensional descriptors. SOM acts as an image classifier, mapping images to neurons in the network. It generates maps where similar images are close in the network and this characteristic is used during retrieval (Fig. 2).

Since working with big networks can reduce the performance of the classical SOM, many different models have been developed. BSOM [11] is an alternative that modifies the training algorithm, reducing the time required to train the net. ParSOM [12] consists of dividing the network into many sections which are maintained by different processing nodes (Fig. 3). In this model, training and retrieval can be performed in parallel. We propose a new model known as *ParBSOM* that combines both characteristics leading to a considerable improvement in training and retrieval times.

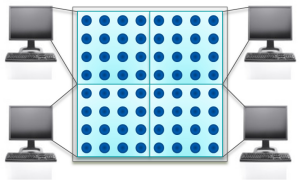

Figure 3: Network is divided and assigned to different nodes

In order to overcome TBIR and CBIR problems, recently the hybrid approach was introduced. CBIR and TBIR produce their own results and then both lists are merged (*late fusion*). One of the merging strategies is known as *refinement* [1], which reorders TBIR results using the results of CBIR. This strategy gives more importance to textual results as nowadays TBIR is a much more advanced area than CBIR.

Finally, we develop a system known as *Envision* that can perform CBIR, TBIR and hybrid queries. It was designed with extensibility and flexibility in mind, so that adding new image descriptors, index structures or merging strategies become easy tasks.

3 Results

Data [size x dimension]	**Map units**	BSOM vs. SOM	ParSOM vs. SOM	ParBSOM vs. BSOM	ParBSOM vs. ParSOM
[5.000 x 250]	500	56%	33%	**34%**	**57%**
	1.000	52%	33%	**35%**	**54%**
[5.000 x 500]	500	55%	31%	**34%**	**58%**
	1.000	50%	31%	**40%**	**57%**
[10.000 x 250]	500	60%	32%	**36%**	**63%**
	1.000	58%	32%	**37%**	**61%**
[10.000 x 500]	500	59%	31%	**38%**	**63%**
	1.000	56%	32%	**40%**	**61%**

Table 1: Improvements in training times for different models (10 epochs of training)

In Table 1, we compared training times for different SOM models: the traditional SOM, BSOM, ParSOM and our proposed model ParBSOM. Data sets of different size and dimension and two processing nodes -for parallel versions- were used in the experiments. As expected, the existing variants (BSOM and ParSOM) reduce training times (above 50% and 30% respectively). In addition, our proposed method improves BSOM by about 40% and also ParSOM by about 60%.

Using the databases described in Section 1, we focused on measuring the quality of the generated maps. We compared ParSOM and ParBSOM with the *Brute Force* algorithm, which consists of performing a linear search through the database. Table 2 shows that ParBSOM loses less than 10% of quality in all databases against the Brute Force method and that ParSOM has a similar behavior.

Image DBs	**Quality loss** ParBSOM vs. Brute Force	**Quality loss** ParSOM vs. Brute Force	**Quality loss** ParSOM vs. ParBSOM
ZuBuD	**0.46%**	1.12%	0.66%
UCID	**8.1%**	10.89%	3.04%
UK Bench	**9.07%**	9.94%	0.97%

Table 2: Quality loss in terms of F-Measure

In spite of losing some quality, ParBSOM considerably improves retrieval times (more than 90%) compared to the Brute Force algorithm, as can be observed in Table 3.

Image DBs	**Time** Brute Force	**Time** ParBSOM	**Improvement** ParBSOM vs. Brute Force
ZuBuD	3.43 ms.	0.27 ms.	**92%**
UCID	4.58 ms.	0.32 ms.	**93%**
UK Bench	40.63 ms.	1.68 ms.	**96%**

Table 3: Time required to retrieve an image from the database

Finally, we applied the studied methods (HSV color histograms and ParBSOM) to a hybrid system which uses the refinement strategy (Table 4). As expected, metrics which are not sensitive to image rankings (Precision, Recall and F-Measure) show no changes as refinement alters TBIR rankings without modifying the results set. MAP and Precision in the first 10 and 20 results show an improvement between 10% and 20%.

Metric	**TBIR**	**Hybrid**	Improvement
MAP	14.94	16.59	**9.95%**
Precision	5.35	5.35	0%
Recall	49.27	49.27	0%
F-Measure	8.26	8.26	0%
Prec(10)	22.33	27.83	**19.76%**
Prec(20)	18.33	22.08	**16.98%**

Table 4: Different retrieval methods for ImageCLEFphoto 2007

4 Discussion and Conclusion

We have studied several techniques applied to VIR. First, we focused on color histograms, comparing their performance in the RGB and HSV space. We have proposed a scoring function for color histograms in order to eliminate irrelevant images from the results list.

Then, we have investigated how SOM can be used as an index in CBIR. We have introduced a new SOM model (ParBSOM) that improves BSOM's training time by about 40% and also ParSOM's training time by about 60% and proposed to use it in CBIR.

We have studied hybrid techniques and observed that the refinement strategy can actually improve textual results by using visual features.

Finally, we have developed a research system known as *Envision* that implements all the studied methods.

Despite the fact that VIR has been one of the most active areas, there are many open issues that still need to be addressed. In the future we intend to investigate image descriptors that combine color with other interesting features such as shape or texture. We also plan to focus on developing new hybrid techniques to combine textual and visual results.

References

[1] M. Grubinger, *Analysis and Evaluation of Visual Information Systems Performance*. PhD thesis, School of Computer Science and Mathematics, Faculty of Health, Engineering and Science, Victoria University, Melbourne, Australia, 2007.

[2] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11–32, 1991.

[3] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

[4] H. Shao, T. Svoboda, and L. van Gool, "ZuBuD — Zurich Buildings Database for Image Based Recognition," tech. rep., Swiss Federal Institute of Technology, Switzerland, 2003.

[5] G. Schaefer and M. Stich, "UCID - An Uncompressed Colour Image Database," in *Storage and Retrieval Methods and Applications for Multimedia 2004*, vol. 5307 of *Proceedings of SPIE*, pp. 472–480, 2004.

[6] D. Nistér and H. Stewénius, "Scalable Recognition with a Vocabulary Tree," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 2161–2168, 2006.

[7] M. Grubinger, P. Clough, A. Hanbury, and H. Müller, "Overview of the ImageCLEFphoto 2007 Photographic Retrieval Task," *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Revised Selected Papers*, pp. 433–444, 2008.

[8] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *ICIP '95: Proceedings of the 1995 International Conference on Image Processing*, vol. 3, (Washington, DC, USA), p. 3528, IEEE Computer Society, 1995.

[9] O. Jonsgård, "Improvements on colour histogram-based CBIR," Master's thesis, Gjøvik University College, Stockholm, Norway, 2005.

[10] C. Böhm, S. Berchtold, and D. A. Keim, "Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases," *ACM Comput. Surv.*, vol. 33, pp. 322–373, September 2001.

[11] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 3 ed., 2001.

[12] P. Tomsich, A. Rauber, and D. Merkl, "parSOM: Using parallelism to overcome memory latency in self-organizing neural networks," in *High Performance Computing and Networking*, pp. 61–5, Society Press, 2000.